Appendix 1: Simulation study comparing eight different methods to estimate adjusted risk ratios (as supplied by the authors)

Methods

To evaluate the performance of different models to estimate risk ratios we conducted simulation studies. The simulated datasets contained a dichotomous exposure X, a dichotomous outcome Y and a continuous confounder Z. The confounder was uniformly distributed between 0 and 1. The exposure was drawn from a binomial distribution with success probability $Pr(X=1|Z) = 0.5 * \exp(\alpha(Z-0.5))$, where the parameter α indicates the effect of the confounder on the exposure. In this way the proportion exposed is approximately 50%. Outcomes were generated by a log-binomial model: $Pr(Y=1|\ Z,X) = \exp(\beta_0 + \beta_1(X-0.5) + \beta_2(Z-0.5))$.

In the simulations, the following parameters were varied: incidence of outcome (5–45% in steps of 5%, with $\beta_0=\log(\text{incidence})$), strength of exposure effect (risk ratio ($\exp(\beta_1)$ of 1.5, and 2.5), and sample size (250, 500, and 1000). Also the strength of confounding was varied. Both the probability of being exposed and the probability of having the outcome was either two ("weak" confounding, $\exp(\alpha) = \exp(\beta_2)=2$) or four ("moderate" confounding, $\exp(\alpha) = \exp(\beta_2)=4$) times higher for those with confounder value 1 as compared to those with confounder value 0. As a result, the crude risk ratio for the case of "weak" confounding was on average 1.66 for the situation of risk ratio=1.5 and 2.79 for the situation of risk ratio=2.5. In case of "moderate" confounding, the crude risk ratios were 2.18 for the situation of risk ratio=1.5 and 3.67 for the situation of risk ratio=2.5. All simulations were carried out in R for Windows, version 2.10 (1).

For every scenario, 1000 datasets were simulated. The eight methods to estimate a risk ratio as described in the full paper as well as logistic regression were applied to all datasets. Since the Mantel–Haenszel method can only be applied in the presence of categorical confounders, we categorized the continuous confounder into quintiles. Estimated regression coefficients for the effect of exposure on the outcome were averaged over 1000 simulations and compared with the "true" effect, i.e., either RR=1.5 or RR=2.5. Also, the estimated variances of the estimator were averaged. These were compared with the empirical variance of the estimator, which was derived from the distribution of regression coefficients. Furthermore, the coverage was calculated, i.e. the proportion of simulations in which the "true" effect was included in the estimated 95% confidence interval of the regression coefficient, which is ideally 95%. Finally, the empirical power, i.e., the proportion of simulations in which the null hypothesis was rejected, was calculated.

Appendix to: Knol MJ, Le Cessie S, Algra A, et al. Overestimation of risk ratios by odds ratios in trials and cohort studies: alternatives for logistic regression.
CMAJ 2011. DOI:10.1503/cmaj.101715.

Log–binomial regression does not always converge, i.e., the model cannot find a solution and therefore no effect estimate is given, and Poisson regression may yield individual predicted probabilities above 1. Therefore, we assessed whether and when this happened in our simulations.

Although it is essentially not quite different to consider one confounder that is strongly related to both exposure and outcome or multiple confounders that have less strong associations with exposure and outcome (2), we conducted some additional simulations in which we included 10 confounders instead of one confounder. We considered a situation in which all 10 confounders were independent and uniformly distributed (between 0 and 1) and a situation in which all 10 confounders were independent and dichotomous (prevalence 50%). It seems unrealistic to think that the association between each of these 10 confounders and the outcome is as strong as the association between the single confounder and the outcome. Therefore, we reduced the associations between each of the 10 confounders and the outcome by a factor 10 (e.g., the strength of the association between confounder and outcome was set at RR = 1.15 instead of RR = 4, or logRR = 0.14 instead of logRR = 1.4), and the associations between the confounders and exposure likewise. The results of the simulations were similar to the simulations with a single confounder as described below.

Results

As the results of the simulations were similar for the sample sizes that were simulated, we only present the results for a sample size of 1000. Figure 1 presents the results of the simulations with a weak confounder (small exposure effect: risk $\text{ratio}_{crude} = 1.7$ and risk $\text{ratio}_{adjusted} = 1.5$ (Figure 1A, B, C); and large exposure effect: risk $\text{ratio}_{crude} = 2.8$ and risk $\text{ratio}_{adjusted} = 2.5$ (Figure 1D, E, F)). For the situation of a small exposure effect, incidences up to 45% could be simulated; for a large exposure effect, incidences up to 40% could be simulated. As expected, the odds ratio obtained with logistic regression overestimated the risk ratio (Figure 1A). This overestimation increased with increasing incidence of the outcome, and with increasing exposure effect (Figure 1D). Methods to estimate a risk ratio yielded unbiased estimates, except for the method by Zhang and Yu, which slightly overestimated the risk ratio with increasing incidence. When comparing the estimated standard errors with the empirical standard errors, Figure 1B shows that ordinary Poisson regression and the doubling of cases method give too large standard errors. The robust standard errors for these two methods are close to the empirical standard errors (ratio estimated/empirical standard errors is close to 1). In addition, Figure 1B shows that the calculated standard errors for the Zhang and Yu method are too small. The pattern of the standard errors is similar for smaller and larger exposure effects (Figure 1E). Due to bias of the effect estimate and a too small standard error, the coverage of the 95% confidence interval of the Zhang and Yu method is lower than 95% (Figure 1C and F) and decreases with increasing incidences. Also, logistic regression has lower coverage than 95%, especially when the exposure effect is large (Figure 1F), and coverage decreases with increasing incidence of the outcome. As ordinary Poisson regression and the doubling of cases method gave too large standard errors, the coverage is too high. When using robust standard errors for these methods, coverage is good, i.e., 95%. Log–binomial regression, the Mantel–Haenszel risk ratio and the method proposed by Austin yielded unbiased estimates of the risk ratio and good coverage.

Figure 2 presents the results of the simulations when there is moderate confounding (small exposure effect: risk $\text{ratio}_{\text{crude}}$ = 2.2 and risk $\text{ratio}_{\text{adjusted}}$ = 1.5 (Figure 2A, B, C); and large exposure effect: risk $\text{ratio}_{\text{crude}}$ = 3.7 and risk $\text{ratio}_{\text{adjusted}}$=2.5 (Figure 2D, E, F)). For the situation of a small exposure effect, incidences up to 35% could be simulated; for a large exposure effect, incidences up to 30% could be simulated. The patterns of bias, standard errors and coverage are similar for moderate and weak confounding. However, the overestimation of the risk ratio by logistic regression is larger for moderate confounding than weak confounding. Furthermore, the Mantel–Haenszel risk ratio slightly overestimates the risk ratio. This is due to some residual confounding because of categorization of the continuous confounder. The method proposed by Austin underestimates the risk ratio in case of a large exposure effect and a large incidence of the outcome, and therefore the coverage of this method is lower than 95%. Similar to the situation with weak confounding, log–binomial regression, Poisson regression with robust standard errors, and the doubling of cases method with robust standard errors gave no bias and good coverage in all situations. Also the empirical standard errors indicating precision were very close for these three methods (data not shown).

The empirical power increased with increasing incidence of the outcome and obviously also with increasing sample size. The Mantel–Haenszel risk ratio, the log–binomial model, the Poisson model with robust standard errors, the doubling of the cases method with robust standard errors, and the method proposed by Austin yielded similar empirical power. For samples of 1000 subjects, this ranged from 30% (incidence of the outcome 5%) to close to 100% when the incidence of the outcome was ≥ 40%. Importantly, the empirical power of logistic regression was similar to the abovementioned methods, meaning that logistic regression can make adequate statistical inference if the incidence of the outcome is not rare. The Poisson model and the doubling of the cases method yielded lower power, due to their (too) large standard errors. The Zhang and Yu method had considerable lower empirical power.

Log–binomial regression did not converge (even after changing the starting values in the numerical algorithm which estimates the parameters) in the situations with the highest simulated incidence, e.g., log-binomial regression did not converge for the situation with weak confounding, a small exposure effect and an incidence of 45%. There were three situations for which Poisson regression gave probabilities above 1, but the chance of estimating probabilities above 1 was very small in these situations: 1) weak confounding, large exposure effect and incidence of 40% (chance of probabilities above 1 = 0.00005); 2) strong confounding, small exposure effect and incidence of 35% (chance of probabilities above 1 = 0.000001); 3) strong confounding, large exposure effect and incidence of 30% (chance of probabilities above 1 = 0.002). Note that predicted probabilities above 1 will be more common with higher incidences; we only simulated up to an incidence of 45%.

References

(1) R Development Core Team. R: A language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing; 2006. ISBN 3-900051-07-0. URL: http://www.R-project.org

(2) Greenland S. Multiple-bias modelling for analysis of observational data. *Journal of the Royal Statistical Society Series A (Statistics in Society)* 2005;168:267-91.

Figure 1

Bias (Figure A, D), ratio of estimated and empirical standard error (Figure B, E) and coverage of 95% confidence interval (Figure C, F) of the eight methods to estimate a risk ratio and logistic regression for the situation with ***weak*** confounding where the true adjusted risk ratio is 1.5 (Figure A, B, C) or 2.5 (Figure D, E, F). Dotted horizontal lines represent the true or correct values. Mantel-Haenszel method to estimate a risk ratio (Δ); log-binomial regression (+); ordinary Poisson regression (□); Poisson regression with robust standard errors (■); method proposed by Zhang and Yu (x); doubling of cases method proposed by Miettinen (o); doubling of cases method with robust standard errors (●); method proposed by Austin (*); logistic regression (▲). Note: in Figure A and D, Poisson regression with robust standard errors and the doubling of cases method with robust standard errors are not depicted as these methods give exactly the same exposure effects as ordinary Poisson regression and the doubling of cases method without robust standard errors.

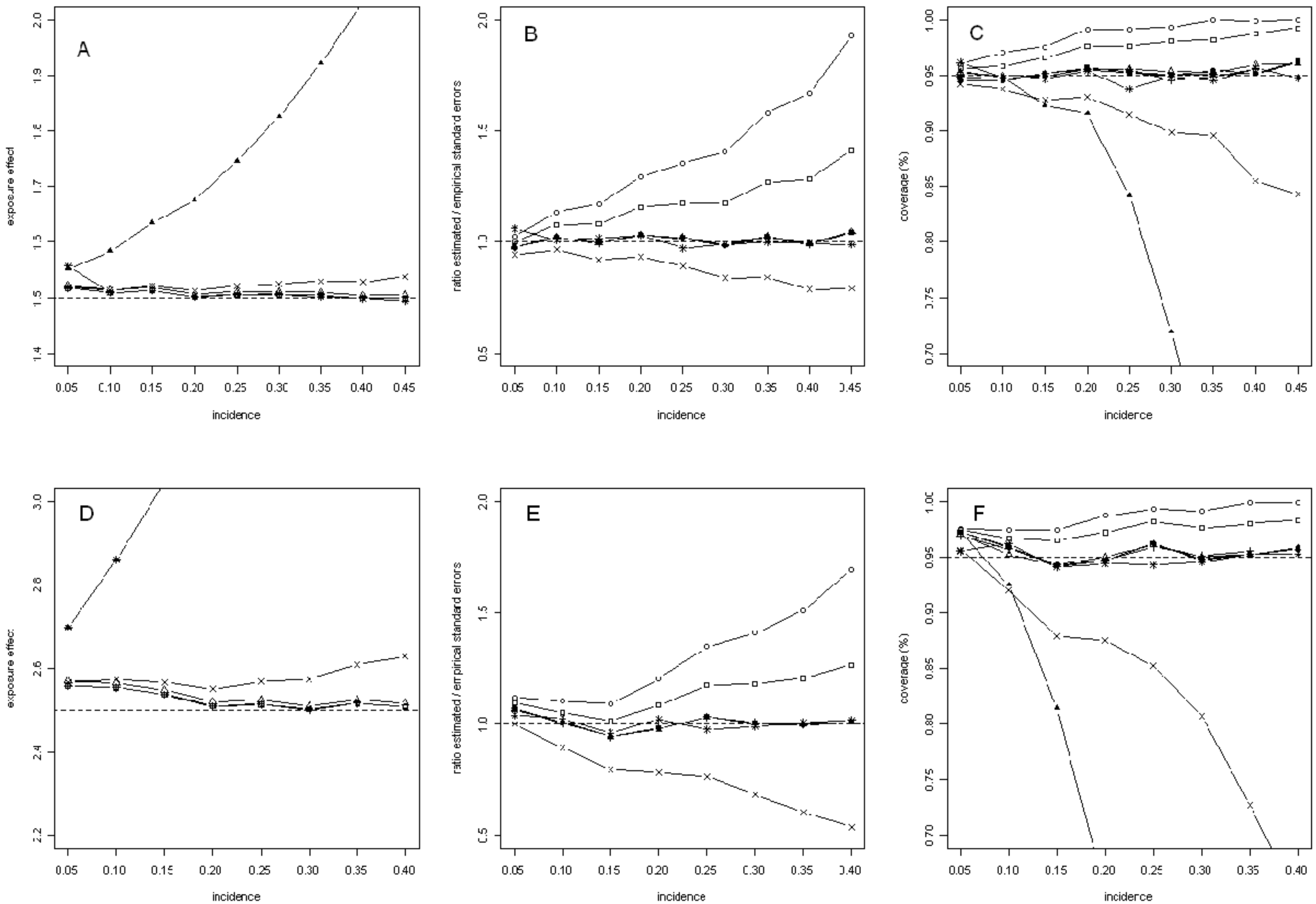
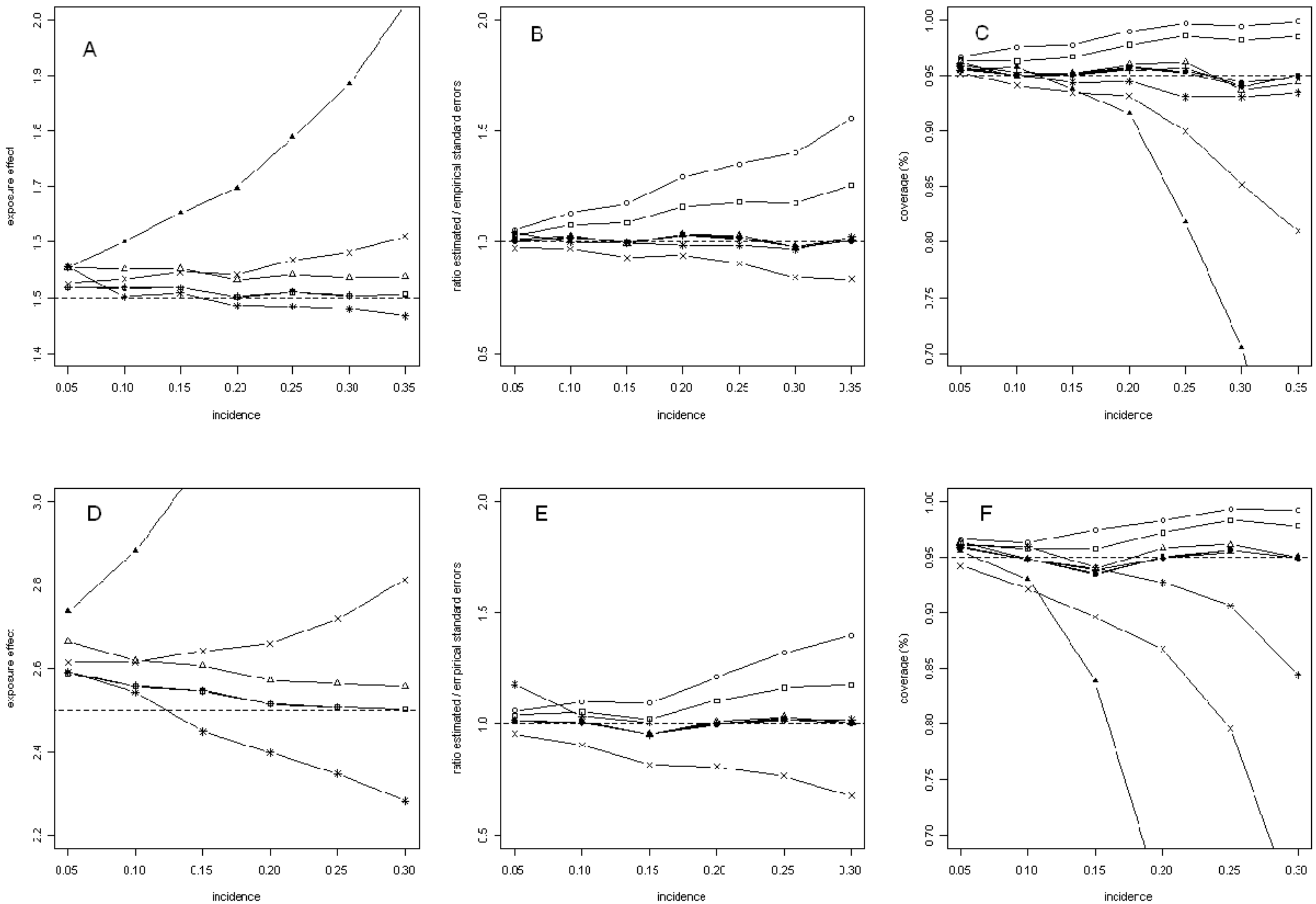

A
B
C
D
E
F
exposure effect
ratio estimated / empirical standard errors
coverage (%)
incidence
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.45

Figure 2

Bias (Figure A, D), ratio of estimated and empirical standard error (Figure B, E) and coverage of 95% confidence interval (Figure C, F) of the eight methods to estimate a risk ratio and logistic regression for the situation with ***moderate*** confounding where the true adjusted risk ratio is 1.5 (Figure A, B, C) or 2.5 (Figure D, E, F). Dotted horizontal lines represent the true or correct values. Mantel-Haenszel method to estimate a risk ratio (Δ); log-binomial regression (+); ordinary Poisson regression (□); Poisson regression with robust standard errors (■); method proposed by Zhang and Yu (x); doubling of cases method proposed by Miettinen (o); doubling of cases method with robust standard errors (●); method proposed by Austin (*); logistic regression (▲). Note: in Figure A and D, Poisson regression with robust standard errors and the doubling of cases method with robust standard errors are not depicted as these methods give exactly the same exposure effects as ordinary Poisson regression and the doubling of cases method without robust standard errors.


A
B
C
D
E
F
exposure effect
ratio estimated / empirical standard errors
coverage (%)
incidence